Course: MMDS

Lab 03

MapReduce

Fill answers of the questions below in the given tables.
Your screenshots must contain commands for required operations.

Question 1:

Given a tsv file WHO-COVID-19-20210601-213841.tsv which is corresponding to the WHO Coronavirus (COVID-19) Dashboard.

Students are required to create a folder, named **lab03**, in HDFS and then copy the tsv to **lab03/input/**

Take a screenshot to show the content of **lab03/input/** in HDFS

My screenshot

```
pkucpkam@PhucPham:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -mkdir /lab03
pkucpkam@PhucPham:~/Desktop/hadoop-3.2.1$ bin/hdfs dfs -mkdir /lab03/input
pkucpkam@PhucPham:~/Desktop/hadoop-3.2.1$ hdfs dfs -put ~/Downloads/WHO-COVID-19-202
10601-213841.tsv /lab03/input/
2025-02-12 00:30:10,825 INFO sasl.SaslDataTransferClient: SASL encryption trust chec
k: localHostTrusted = false, remoteHostTrusted = false
pkucpkam@PhucPham:~/Desktop/hadoop-3.2.1$ hdfs dfs -ls /user/pkucpkam/lab03/input/
ls: `/user/pkucpkam/lab03/input/': No such file or directory
pkucpkam@PhucPham:~/Desktop/hadoop-3.2.1$ hdfs dfs -ls /lab03/input/
Found 1 items
-rw-r--r--   2 pkucpkam supergroup      28907 2025-02-12 00:30 /lab03/input/WHO-COVI
D-19-20210601-213841.tsv
pkucpkam@PhucPham:~/Desktop/hadoop-3.2.1$
```

Question 2:

Create one and only one java file, named **ASEANCaseCount.java**, to run a MapReduce job that counts the number of cumulative total cases among ASEAN countries (*South-East Asia Region in the given data table*).
The output of the MapReduce job is located in **lab03/output-java/**.
Submit the source code file following the instructions in Submission Notice.

Result screen:

```
pkucpkam@PhucPham:~/Desktop/hadoop-3.2.1$ hdfs dfs -cat /lab03/output-java/part-
r-00000
2025-02-22 07:42:04,072 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
South-East Asia 31923614
```

Question 3 (*optional*):

Create a pair of Python files, named **ASEANDeathCountMapper.py** and **ASEANDeathCountReducer.py**, to run a MapReduce job that counts the number of cumulative total deaths among ASEAN countries (*South-East Asia Region in the given data table*).
The output of the MapReduce job is located in **lab03/output-python/**.
Submit the source code files following the instructions in Submission Notice.

Result screen:

```
pkucpkam@PhucPham:~/Desktop/hadoop-3.2.1$ hadoop fs -cat /lab03/output-python/pa
rt-00000
2025-02-22 08:01:46,807 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
South-East Asia 408426
```